Web-based Supplementary Materials for "An Alternative Robust Estimator of Average Treatment Effect in Causal Inference"

by Jianxuan Liu, Yanyuan Ma and Lan Wang

This online supplementary material contains the regularity conditions, the proof of Lemmas and additional numerical results.

S.1 Regularity conditions

(C1) The univariate kernel function $K(\cdot)$ is Lipschitz, symmetric and has compact support. It satisfies

$$\int K(u)du = 1,\ \int u^i K(u)du = 0, 1 \leq i \leq m-1,\ 0 \neq \int u^m K(u)du < \infty.$$

The d-dimensional kernel function is a product of d univariate kernel functions, that is, $K_h(\mathbf{u}) = K(\mathbf{u}/h)/h^d = \Pi_{j=1}^d K_h(u_j) = \Pi_{j=1}^d K(u_j/h)/h^d$ for $\mathbf{u} = (u_1, \ldots, u_d)^{\mathrm{T}}$. Here we abuse the notation and use the same K regardless of the dimension of its argument.

(C2) $E(\mathbf{X} \mid \mathcal{B}^{\mathrm{T}}\mathbf{x})f(\mathcal{B}^{\mathrm{T}}\mathbf{x})$ and $\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x})$ are mth order differentiable and their m-th derivatives, as well as $f(\mathbf{x}^{\mathrm{T}}\mathcal{B})$ are locally Lipschitz-continuous.

(C3) The density functions of $\mathbf{X}$ and $\mathcal{B}^{\mathrm{T}}\mathbf{X}$, denoted respectively by $f_{\mathbf{X}}(\mathbf{x})$ and $f(\mathcal{B}^{\mathrm{T}}\mathbf{x})$, are bounded from below and above. Each entry in the matrices $E\left(\mathbf{X}\mathbf{X}^{\mathrm{T}} \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}\right)$ is locally Lipschitz-continuous and bounded from above as a function of $\mathcal{B}^{\mathrm{T}}\mathbf{x}$.

(C4) The bandwidth $h = O(n^{-\kappa})$ for $1/(4m) < \kappa < 1/(2d)$.

(C5) $\mathrm{vecl}\left(\{\mathbf{x}_i - E(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i - \frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)$ is differentiable w.r.t $\mathrm{vecl}(\mathcal{B})$.

(C6) $E\left\{\frac{\partial}{\partial(\mathrm{vecl}\ \mathcal{B})^{\mathrm{T}}}\mathrm{vecl}\left(\{\mathbf{x}_i - E(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i - \frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)\right\}$ is invertible and bounded.

(C7) $\mathrm{vecl}\left(\{\mathbf{x}_i - E(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i - \frac{\exp\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}'(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)$ is differentiable w.r.t $\mathrm{vecl}(\mathcal{B})$.

(C8) $E\left\{\frac{\partial}{\partial(\mathrm{vecl}\ \mathcal{B})^{\mathrm{T}}}\mathrm{vecl}\left(\{\mathbf{x}_i - E(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i - \frac{\exp\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}'(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)\right\}$ is invertible and bounded.

S.2 Proof of Lemma 1

We first summarize and prove a result that will be used in the proof of Lemma 1

Lemma S.2.1.

$$\frac{E\{\mathbf{X}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) \mid \mathbf{x}_i\}}{E\{K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) \mid \mathbf{x}_i\}} - E(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) = O(h^m).$$

Proof: Let $\mathbf{z} = \frac{\mathcal{B}^{\mathrm{T}}\mathbf{x}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i}{h}$, then $\mathcal{B}^{\mathrm{T}}\mathbf{x}_j = \mathbf{z}h + \mathcal{B}^{\mathrm{T}}\mathbf{x}_i$. Let $f(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)$ be the probability density function of $\mathcal{B}^{\mathrm{T}}\mathbf{X}_j$. Note that we have

$$\begin{aligned} E\{K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) \mid \mathbf{x}_i\} &= \int h^{-d}K\{(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)/h\}f(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j)d\mu(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j) \\ &= \int f(\mathbf{z}h + \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)K(\mathbf{z})d\mu(\mathbf{z}) \\ &= f(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i) + O(h^m). \end{aligned}$$

Similarity,

$$\begin{aligned} E\{\mathbf{X}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) \mid \mathbf{x}_i\} &= E\{E(\mathbf{X}_j \mid \mathcal{B}^{\mathrm{T}}\mathbf{X}_j)K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) \mid \mathbf{x}_i\} \\ &= \int E(\mathbf{X}_j \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_j)h^{-d}K\{(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)/h\}f(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j)d\mu(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j) \\ &= \int E(\mathbf{X}_j \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i + \mathbf{z}h)f(\mathbf{z}h + \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)K(\mathbf{z})d\mu(\mathbf{z}) \\ &= E(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)f(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i) + O(h^m) \end{aligned}$$

Thus we have

$$\frac{E\{\mathbf{X}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) \mid \mathbf{x}_i\}}{E\{K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) \mid \mathbf{x}_i\}} - E(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) = O(h^m).$$

$\square$

Proof of Lemma 1: Standard nonparametric kernel estimation results yield

$$\begin{aligned} & n^{-1}\sum_{j=1}^{n} K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) - E\{K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\} \\ = \;& n^{-1}\sum_{j=1}^{n} K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) - f(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i) + O(h^m) \\ = \;& O_p\{(nh^d)^{-1/2}\} + O(h^m) \end{aligned}$$

and

$$\begin{aligned}
& n^{-1}\sum_{j=1}^{n}\mathbf{X}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j-\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)-E\{\mathbf{X}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j-\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\} \\
= & n^{-1}\sum_{j=1}^{n}\mathbf{X}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j-\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)-E(\mathbf{X}_i\mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)f(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)+O(h^m) \\
= & O_p\{(nh^d)^{-1/2}\}+O(h^m).
\end{aligned}$$

From the estimating equation, the locally efficient estimator $\widehat{\mathcal{B}}_1$ satisfies

$$\begin{aligned}
\mathbf{0} = & n^{-1/2}\sum_{i=1}^{n}\mathrm{vecl}\left(\{\mathbf{x}_i-\widehat{E}(\mathbf{X}_i\mid \widehat{\mathcal{B}}_1^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i-\frac{\exp\{\eta^*(\widehat{\mathcal{B}}_1^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\widehat{\mathcal{B}}_1^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\widehat{\mathcal{B}}_1^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right) \\
= & n^{-1/2}\sum_{i=1}^{n}\mathrm{vecl}\left(\{\mathbf{x}_i-\widehat{E}(\mathbf{X}_i\mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i-\frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right) \\
& +n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial(\mathrm{vecl}\ \mathcal{B})^{\mathrm{T}}}\mathrm{vecl}\left(\{\mathbf{x}_i-\widehat{E}(\mathbf{X}_i\mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i-\frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)\Bigg|_{\mathcal{B}=\mathcal{B}^*} \\
& \times\sqrt{n}\mathrm{vecl}(\widehat{\mathcal{B}}_1-\mathcal{B}) \\
= & n^{-1/2}\sum_{i=1}^{n}\mathrm{vecl}\left(\{\mathbf{x}_i-\widehat{E}(\mathbf{X}_i\mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i-\frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right) \\
& +E\left\{\frac{\partial}{\partial(\mathrm{vecl}\ \mathcal{B})^{\mathrm{T}}}\mathrm{vecl}\left(\{\mathbf{X}_i-\widehat{E}(\mathbf{X}_i\mid \mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}\left[T_i-\frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)^{\mathrm{T}}\right)\right\} \\
& \times\sqrt{n}\mathrm{vecl}(\widehat{\mathcal{B}}_1-\mathcal{B})+o_p(1),
\end{aligned}$$

where $\mathcal{B}^*$ is on the line connecting $\mathcal{B}$ and $\widehat{\mathcal{B}}_1$, and we used a Taylor expansion at $\mathcal{B}$ to obtain the second equality.

Using the kernel estimation form of $\widehat{E}(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)$, we have

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\left(\{\mathbf{x}_i - \widehat{E}(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i - \frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right) \\
= \;& n^{-1/2}\sum_{i=1}^{n}\left(\left\{\mathbf{x}_i - \frac{n^{-1}\sum_{j=1}^{n}\mathbf{x}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)}{n^{-1}\sum_{j=1}^{n} K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)}\right\}\left[T_i - \frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right) \\
= \;& n^{-1/2}\sum_{i=1}^{n}\left(\left[\mathbf{x}_i - \frac{E\{\mathbf{X}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{E\{K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\left[T_i - \frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right) \\
& -n^{-1/2}\sum_{i=1}^{n}\left(\left[\frac{n^{-1}\sum_{j=1}^{n}\mathbf{x}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) - E\{\mathbf{X}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{E\{K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\right. \\
& \left.\times\left[T_i - \frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right) \\
& +n^{-1/2}\sum_{i=1}^{n}\left\{\left(\frac{E\{\mathbf{X}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[n^{-1}\sum_{j=1}^{n} K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) - E\{K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\right.}{[E\{K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}]^2}\right.\right. \\
& \left.\times\left[T_i - \frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right\} + O_p(n^{1/2}h^{2m} + n^{-1/2}h^{-d}) \\
= \;& n^{-1/2}\sum_{i=1}^{n}\left(\{\mathbf{x}_i - E(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i - \frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right) \\
& -n^{-1/2}\sum_{i=1}^{n}\left(\left[\frac{n^{-1}\sum_{j=1}^{n}\{\mathbf{x}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) - E(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{f(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)}\right]\right. \\
& \left.\times\left[T_i - \frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right) + O_p(n^{1/2}h^{m} + n^{1/2}h^{2m} + n^{-1/2}h^{-d}),
\end{aligned}
$$

where the last equality is due to Lemma S.2.1. Note that also from Lemma S.2.1, we have

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\left(\left[\frac{E\{\mathbf{X}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) \mid \mathbf{x}_i\} - E(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)E\{K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_j - \mathcal{B}^{\mathrm{T}}\mathbf{x}_i) \mid \mathbf{x}_i\}}{f(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)}\right]\right. \\
& \left.\times\left[T_i - \frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right) \\
= \;& O(n^{1/2}h^{m}).
\end{aligned}
$$

Using the U-statistic technique and applying the above result, we obtain

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\left(\left[\frac{n^{-1}\sum_{j=1}^{n}\{\mathbf{x}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j-\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)-E(\mathbf{X}_i\mid\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j-\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{f(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)}\right]\right.\\
& \left.\times\left[T_i-\frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)\\
= \;& n^{-1/2}\sum_{j=1}^{n}E\left(\left\{\frac{\mathbf{x}_j K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j-\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)-E(\mathbf{X}_i\mid\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j-\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)}{f(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)}\right\}\right.\\
& \left.\times\left[T_i-\frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)^{\mathrm{T}}\mid\mathbf{x}_j\right)+O(n^{1/2}h^m)+o_p(1)\\
= \;& n^{-1/2}\sum_{j=1}^{n}E\left(K_h(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j-\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\left\{\frac{\mathbf{x}_j-E(\mathbf{X}_i\mid\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)}{f(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)}\right\}\right.\\
& \left.\times\left[\frac{\exp\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}}{1+\exp\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}}-\frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)^{\mathrm{T}}\mid\mathbf{x}_j\right)+O(n^{1/2}h^m)+o_p(1)\\
= \;& n^{-1/2}\sum_{j=1}^{n}\left\{\mathbf{x}_j-E(\mathbf{X}_j\mid\mathcal{B}^{\mathrm{T}}\mathbf{x}_j)\right\}\left[\frac{\exp\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j)\}}{1+\exp\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j)\}}-\frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_j)^{\mathrm{T}}\\
& +O(n^{1/2}h^m)+o_p(1).
\end{aligned}
$$

Combining the above results, we obtain

$$
\begin{aligned}
\mathbf{0} \;=\;& n^{-1/2}\sum_{i=1}^{n}\mathrm{vecl}\left(\left\{\mathbf{x}_i-E(\mathbf{X}_i\mid\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\right\}\left[T_i-\frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)\\
& -n^{-1/2}\sum_{i=1}^{n}\mathrm{vecl}\left(\left\{\mathbf{x}_i-E(\mathbf{X}_i\mid\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\right\}\left[\frac{\exp\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}-\frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)\\
& +E\left\{\frac{\partial}{\partial(\mathrm{vecl}\;\mathcal{B})^{\mathrm{T}}}\mathrm{vecl}\left(\{\mathbf{x}_i-E(\mathbf{X}_i\mid\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i-\frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)\right\}\\
& \times\sqrt{n}\,\mathrm{vecl}(\widehat{\mathcal{B}}_1-\mathcal{B})+o_p(1)\\
=\;& n^{-1/2}\sum_{i=1}^{n}\mathrm{vecl}\left(\left\{\mathbf{x}_i-E(\mathbf{X}_i\mid\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\right\}\left[T_i-\frac{\exp\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}^{\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)\\
& +E\left\{\frac{\partial}{\partial(\mathrm{vecl}\;\mathcal{B})^{\mathrm{T}}}\mathrm{vecl}\left(\{\mathbf{X}_i-E(\mathbf{X}_i\mid\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}\left[T_i-\frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)^{\mathrm{T}}\right)\right\}\\
& \times\sqrt{n}\,\mathrm{vecl}(\widehat{\mathcal{B}}_1-\mathcal{B})+o_p(1),
\end{aligned}
$$

which leads to the asymptotically normal results in Lemma 1. Further, it is easy to check that when $\eta^* = \eta$,

$$
\begin{aligned}
\mathbf{A} &= E\left\{\frac{\partial}{\partial(\mathrm{vecl}\ \mathcal{B})^{\mathrm{T}}}\mathrm{vecl}\left(\{\mathbf{X}_i - E(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}\left[T_i - \frac{\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}}{1+\exp\{\eta^*(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\}}\right]\boldsymbol{\eta}^{*\prime}(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)^{\mathrm{T}}\right)\right\} \\
&= E\left\{\partial\mathbf{S}_{\mathrm{eff}}(T_i, \mathbf{X}_i, \mathcal{B}^{\mathrm{T}}\mathbf{X}_i, \eta, \boldsymbol{\eta}')/\partial(\mathrm{vecl}\ \mathcal{B})^{\mathrm{T}}\right\} \\
&= -E\left\{\mathbf{S}_{\mathrm{eff}}(T_i, \mathbf{X}_i, \mathcal{B}^{\mathrm{T}}\mathbf{X}_i, \eta, \boldsymbol{\eta}')^{\otimes 2}\right\}
\end{aligned}
$$

while $\mathcal{B} = E\left\{\mathbf{S}_{\mathrm{eff}}(T_i, \mathbf{X}_i, \mathcal{B}^{\mathrm{T}}\mathbf{X}_i, \eta, \boldsymbol{\eta}')^{\otimes 2}\right\}$, thus the estimator is efficient. $\square$

S.3 Proof of Lemma 2

For notational brevity, we write $\mathbf{t}_i = \mathcal{B}^{\mathrm{T}}\mathbf{x}_i, \widetilde{\mathbf{t}}_i = \widetilde{\mathcal{B}}^{\mathrm{T}}\mathbf{x}_i$. We also write $H(u) = \exp(u)/\{1+\exp(u)\}$.

We first demonstrate the consistency of $\widehat{b}_0$ and $\widehat{\mathbf{b}}_1$. Denote $b_0 = \lim_{n\to\infty}\widehat{b}_0$ and $\mathbf{b}_1 = \lim_{n\to\infty}\widehat{\mathbf{b}}_1$ with probability 1. With $\mathcal{B}$ replaced by $\widetilde{\mathcal{B}}$, (2.7) can be equivalently written as

$$
\begin{aligned}
0 &= n^{-1}\sum_{i=1}^{n}\left[T_i - H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(\widetilde{\mathbf{t}}_i - \mathbf{t})\}\right]K_h(\widetilde{\mathbf{t}}_i - \mathbf{t}) \\
&= n^{-1}\sum_{i=1}^{n}\left[y_i - H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(\mathbf{t}_i - \mathbf{t})\}\right]K_h(\mathbf{T}_i - \mathbf{t}) + O_p(n^{-1/2}) \qquad \text{(S.1)}
\end{aligned}
$$

at any fixed h. Let $n \to \infty$, we then obtain

$$
\begin{aligned}
0 &= E\left[T_i - H\{b_0 + \mathbf{b}_1^{\mathrm{T}}(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})\} \mid \mathcal{B}^{\mathrm{T}}\mathbf{X}_i = \mathbf{t}\right]f(\mathbf{t}) + o(1) \\
&= [H\{\eta(\mathbf{t})\} - H(b_0)]\,f(\mathbf{t}) + o(1).
\end{aligned}
$$

Because both $f(\mathbf{t})$ and H' are positive, we obtain $b_0 = \eta(\mathbf{t})$, i.e. $\widehat{b}_0 \to \eta(\mathbf{t})$ in probability when $n \to \infty$. Similarly, (2.8) can be equivalently written as

$$\begin{aligned}
\mathbf{0} &= n^{-1}\sum_{i=1}^{n}\left[T_i - H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(\widetilde{\mathbf{t}}_i - \mathbf{t})\}\right](\widetilde{\mathbf{t}}_i - \mathbf{t})^{m-1}K_h(\widetilde{\mathbf{t}}_i - \mathbf{t}) \\
&= n^{-1}\sum_{i=1}^{n}\left[T_i - H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(\mathbf{t}_i - \mathbf{t})\}\right](\mathbf{t}_i - \mathbf{t})^{m-1}K_h(\mathbf{t}_i - \mathbf{t}) + O_p(n^{-1/2}) \qquad \text{(S.2)}
\end{aligned}$$

at any fixed h. Let $n \to \infty$, then

$$n^{-1}\sum_{i=1}^{n}\left[T_i - H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(\mathbf{t}_i - \mathbf{t})\}\right](\mathbf{t}_i - \mathbf{t})^{m-1}K_h(\mathbf{t}_i - \mathbf{t}) \to \mathbf{0}.$$

in probability for any fixed h. On the other hand,

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\left[T_i - H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(\mathbf{t}_i - \mathbf{t})\}\right](\mathbf{t}_i - \mathbf{t})^{m-1}K_h(\mathbf{t}_i - \mathbf{t}) \\
\to\ & E\left(\left[T_i - H\{b_0 + \mathbf{b}_1^{\mathrm{T}}(\mathbf{T}_i - \mathbf{t})\}\right](\mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})^{m-1}K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})\right) \\
=\ & E\left(\left[H\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\} - H\{b_0 + \mathbf{b}_1^{\mathrm{T}}(\mathbf{T}_i - \mathbf{t})\}\right](\mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})^{m-1}K_h(\mathbf{T}_i - \mathbf{t})\right) \\
=\ & \left([H'\{\eta(\mathbf{t})\}\eta'(\mathbf{t}) - H'(b_0)\mathbf{b}_1]\, f(\mathbf{t}) + [H\{\eta(\mathbf{t})\} - H(b_0)]\, f'(\mathbf{t})\right)h^m \int \mathbf{u}^m K(\mathbf{u})d\mathbf{u} + O(h^{m+2}) \\
=\ & H'\{\eta(\mathbf{t})\}\{\eta'(\mathbf{t}) - \mathbf{b}_1\}f(\mathbf{t})h^m \int \mathbf{u}^m K(\mathbf{u})d\mathbf{u} + O(h^{m+2})
\end{aligned}$$

at any fixed sufficiently small h. Thus, $\mathbf{b}_1 = \eta'(\mathbf{t})$, i.e. $\widehat{\mathbf{b}}_1 \to \eta'(\mathbf{t})$ in probability when $n \to \infty$.

We now analyze the bias property of $\widehat{b}_0$ and $\widehat{\mathbf{b}}_1$. Taking expectation of (S.1), we obtain

$$
\begin{aligned}
0 &= E\left(\left[T_i - H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(\ \mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})\}\right] K_h(\ \mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})\right) + O(n^{-1/2}) \\
&= E\left(\left[H\{\eta(\ \mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\} - H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(\ \mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})\}\right] K_h(\ \mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})\right) + O(n^{-1/2}) \\
&= \int \left(H\{\eta(\mathbf{t}_i)\} - E\left[H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(\mathbf{t}_i - \mathbf{t})\} \mid \ \mathcal{B}^{\mathrm{T}}\mathbf{X}_i = \mathbf{t}_i\right]\right) K_h(\mathbf{t}_i - \mathbf{t}) f(\mathbf{t}_i) d\mu(\mathbf{t}_i) + O(n^{-1/2}) \\
&= \int \left(H\{\eta(\mathbf{t} + h\mathbf{u})\} - E\left[H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(h\mathbf{u})\} \mid \ \mathcal{B}^{\mathrm{T}}\mathbf{X}_i = \mathbf{t} + h\mathbf{u}\right]\right) K(\mathbf{u}) f(\mathbf{t} + h\mathbf{u}) d\mathbf{u} + O(n^{-1/2}) \\
&= \left[H\{\eta(\mathbf{t})\} - E\left\{H(\widehat{b}_0) \mid \ \mathcal{B}^{\mathrm{T}}\mathbf{X}_i = \mathbf{t}\right\}\right] f(\mathbf{t}) + h^m \int \mathbf{u}^m K(\mathbf{u}) d\mathbf{u} \\
&\quad \times \frac{\partial^m}{\partial(h\mathbf{u})^m} \left(H\{\eta(\mathbf{t} + h\mathbf{u})\} - E\left[H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(h\mathbf{u})\} \mid \ \mathcal{B}^{\mathrm{T}}\mathbf{X}_i = \mathbf{t} + h\mathbf{u}\right]\right) f(\mathbf{t} + h\mathbf{u})\Big|_{h\mathbf{u}=\mathbf{0}} \\
&\quad + O(h^{m+2} + n^{-1/2}).
\end{aligned}
$$

We therefore obtain $E\{H(\widehat{b}_0) \mid \ \mathcal{B}^{\mathrm{T}}\mathbf{X}_i = \mathbf{t}\} f(\mathbf{t}) - H\{\eta(\mathbf{t})\} = O(h^m)$, hence further $E\{H(\widehat{b}_0)\} - H\{\eta(\mathbf{t})\} = O(h^m)$ since the change of nonparametric estimator $\widehat{b}_0$ is ignorable when one random observation is replaced by a fixed point. Because of the consistency of $\widehat{b}_0$, we obtain $E\{H(\widehat{b}_0)\} = E\{H(b_0) + H'(b_0)(\widehat{b}_0 - b_0) + O_p(\widehat{b}_0 - b_0)^2\} = H\{\eta(\mathbf{t})\} + O(h^m)$, i.e., $H'(b_0)E(\widehat{b}_0 - b_0) = O(h^m)$. Since $H'(b_0)$ is positive, $E(\widehat{b}_0) - \eta(\mathbf{t}) = O(h^m)$.

Similarly, taking expectation of (S.2), we obtain

$$
\begin{aligned}
\mathbf{0} &= E\left(\left[T_i - H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})\}\right](\mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})^{m-1}K_h(\mathbf{T}_i - \mathbf{t})\right) + O(n^{-1/2}) \\
&= E\left(\left[H\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\} - H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})\}\right](\mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})^{m-1}K_h(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i - \mathbf{t})\right) + O(n^{-1/2}) \\
&= \int \left(H\{\eta(\mathbf{t}_i)\} - E\left[H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(\mathbf{t}_i - \mathbf{t})\} \mid \mathcal{B}^{\mathrm{T}}\mathbf{X}_i = \mathbf{t}_i\right]\right)(\mathbf{t}_i - \mathbf{t})^{m-1}K_h(\mathbf{t}_i - \mathbf{t})f(\mathbf{t}_i)d\mu(\mathbf{t}_i) + O(n^{-1/2}) \\
&= \int \left(H\{\eta(\mathbf{t} + h\mathbf{u})\} - E\left[H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(h\mathbf{u})\} \mid \mathcal{B}^{\mathrm{T}}\mathbf{X}_i = \mathbf{t} + h\mathbf{u}\right]\right)(h\mathbf{u})^{m-1}K_h(\mathbf{u})f(\mathbf{t} + h\mathbf{u})d\mathbf{u} + O(n^{-1/2}) \\
&= \frac{\partial}{\partial(h\mathbf{u})}\left(H\{\eta(\mathbf{t} + h\mathbf{u})\} - E\left[H\{\widehat{b}_0 + \widehat{\mathbf{b}}_1^{\mathrm{T}}(h\mathbf{u})\} \mid \mathcal{B}^{\mathrm{T}}\mathbf{X}_i = \mathbf{t} + h\mathbf{u}\right]\right)f(\mathbf{t} + h\mathbf{u})\Big|_{h\mathbf{u}=\mathbf{0}} \\
&\quad \times h^m \int \mathbf{u}^m K(\mathbf{u})d\mathbf{u} + O(h^{m+2} + n^{-1/2}) \\
&= \left(\left[H\{\eta(\mathbf{t})\} - E\left\{H(\widehat{b}_0) \mid \mathcal{B}^{\mathrm{T}}\mathbf{X}_i = \mathbf{t}\right\}\right]f'(\mathbf{t})\right. \\
&\quad \left. + \left[H'\{\eta(\mathbf{t})\}\eta'(\mathbf{t}) - E\left\{H'(\widehat{b}_0)\widehat{\mathbf{b}}_1 \mid \mathcal{B}^{\mathrm{T}}\mathbf{X}_i = \mathbf{t}\right\} - E\left\{H(\widehat{b}_0)\mathbf{S}(\widehat{b}_0, \widehat{\mathbf{b}}_1, \mathbf{t}) \mid \mathcal{B}^{\mathrm{T}}\mathbf{X}_i = \mathbf{t}\right\}\right]f(\mathbf{t})\right) \\
&\quad \times h^m \int \mathbf{u}^m K(\mathbf{u})d\mathbf{u} + O(h^{m+2} + n^{-1/2}) \\
&= \left(\left[H\{\eta(\mathbf{t})\} - E\left\{H(\widehat{b}_0)\right\}\right]f'(\mathbf{t}) + \left[H'\{\eta(\mathbf{t})\}\eta'(\mathbf{t}) - E\left\{H'(\widehat{b}_0)\widehat{\mathbf{b}}_1\right\}\right]f(\mathbf{t})\right) \\
&\quad \times h^m \int \mathbf{u}^m K(\mathbf{u})d\mathbf{u} + O(h^{m+2} + n^{-1/2}) \\
&= \left(H'\{\eta(\mathbf{t})\}\eta'(\mathbf{t}) - E\left[\{H'(b_0) + H''(b_0)(\widehat{b}_0 - b_0)\}\{\mathbf{b}_1 + (\widehat{\mathbf{b}}_1 - \mathbf{b}_1)\}\right]\right)f(\mathbf{t})h^m \int \mathbf{u}^m K(\mathbf{u})d\mathbf{u} \\
&\quad + O(h^{m+2} + n^{-1/2}) \\
&= -H'(b_0)E(\widehat{\mathbf{b}}_1 - \mathbf{b}_1)f(\mathbf{t})h^m \int \mathbf{u}^m K(\mathbf{u})d\mathbf{u} + O(h^{m+2} + n^{-1/2}).
\end{aligned}
$$

Hence $E(\widehat{\mathbf{b}}_1) = \eta'(\mathbf{t}) + O(h^2)$.

We now analyze the asymptotic variance property of $\widehat{b}_0$ and $\widehat{\mathbf{b}}_1$. Let $\mathbf{b} = (b_0, \mathbf{b}_1^{\mathrm{T}})^{\mathrm{T}}$, we write (2.7) and (2.8) jointly as

$$
\begin{aligned}
\mathbf{0} &= n^{-1}\sum_{i=1}^{n}\mathbf{R}(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i,\widetilde{\mathbf{t}}_i-\widetilde{\mathbf{t}},\widehat{\mathbf{b}})K_h(\mathbf{t}_i-\mathbf{t})\\
&= n^{-1}\sum_{i=1}^{n}\mathbf{R}(T_i,\mathbf{t}_i-\mathbf{t},\widehat{\mathbf{b}})K_h(\mathbf{t}_i-\mathbf{t})+O_p(n^{-1/2})\\
&= n^{-1}\sum_{i=1}^{n}\mathbf{R}(T_i,\mathbf{t}_i-\mathbf{t},\mathbf{b})K_h(\mathbf{t}_i-\mathbf{t})+n^{-1}\sum_{i=1}^{n}\left.\frac{\partial\mathbf{R}(T_i,\mathbf{t}_i-\mathbf{t},\mathbf{b})}{\partial\mathbf{b}^{\mathrm{T}}}\right|_{\mathbf{b}=\mathbf{b}^*}K_h(\mathbf{t}_i-\mathbf{t})(\widehat{\mathbf{b}}-\mathbf{b})\\
&\quad +O_p(n^{-1/2})\\
&= n^{-1}\sum_{i=1}^{n}\mathbf{R}(T_i,\mathbf{t}_i-\mathbf{t},\mathbf{b})K_h(\mathbf{t}_i-\mathbf{t})+E\left\{\frac{\partial\mathbf{R}(H\{\eta(\mathbf{T}_i)\},\mathbf{T}_i-\mathbf{t},\mathbf{b})}{\partial\mathbf{b}^{\mathrm{T}}}\mid\mathbf{T}_i=\mathbf{t}\right\}\\
&\quad \times\{1+o_p(1)\}(\widehat{\mathbf{b}}-\mathbf{b})+O_p(n^{-1/2}).
\end{aligned}
$$

Following the same derivation as in the bias analysis, we can obtain

$$
\begin{aligned}
&E\left\{\frac{\partial\mathbf{R}(H\{\eta(\mathcal{B}^{\mathrm{T}}\mathbf{X}_i)\},\ \mathcal{B}^{\mathrm{T}}\mathbf{X}_i-\mathbf{t},\mathbf{b})}{\partial\mathbf{b}^{\mathrm{T}}}\mid\ \mathcal{B}^{\mathrm{T}}\mathbf{X}_i=\mathbf{t}\right\}\\
=&\left\{\begin{array}{cc}
\frac{-\exp\{\eta(\mathbf{t})\}f(\mathbf{t})}{[1+\exp\{\eta(\mathbf{t})\}]^2} & \frac{1}{(m-1)!}\left(\frac{-\exp\{\eta(\mathbf{t})\}f(\mathbf{t})}{[1+\exp\{\eta(\mathbf{t})\}]^2}\right)^{(m-1)}h^m\int\mathbf{u}^mK(\mathbf{u})d\mathbf{u}\\
\left(\frac{-\exp\{\eta(\mathbf{t})\}f(\mathbf{t})}{[1+\exp\{\eta(\mathbf{t})\}]^2}\right)'h^m\int\mathbf{u}^mK(\mathbf{u})d\mathbf{u} & \frac{-\exp\{\eta(\mathbf{t})\}f(\mathbf{t})}{[1+\exp\{\eta(\mathbf{t})\}]^2}h^m\int\mathbf{u}^mK(\mathbf{u})d\mathbf{u}
\end{array}\right\}.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&\operatorname{cov}\left\{n^{-1}\sum_{i=1}^{n}\mathbf{R}(T_i,\mathbf{t}_i-\mathbf{t},\mathbf{b})K_h(\mathbf{t}_i-\mathbf{t})\right\}\\
=\ & n^{-1}E\left\{\mathbf{R}(T_i,\mathbf{t}_i-\mathbf{t},\mathbf{b})K_h(\mathbf{t}_i-\mathbf{t})\right\}^{\otimes 2}-O(h^{2m})\\
=\ & n^{-1}\int E\{\mathbf{R}^{\otimes 2}(T_i,\mathbf{t}_i-\mathbf{t},\mathbf{b})\mid\mathbf{t}_i\}K_h^2(\mathbf{t}_i-\mathbf{t})f(\mathbf{t}_i)d\mathbf{t}_i-O(h^{2m}).
\end{aligned}
$$

We can verify that

$$
\begin{aligned}
& \int E\{\mathbf{R}^{\otimes 2}(T_i, \mathbf{t}_i - \mathbf{t}, \mathbf{b}) \mid \mathbf{t}_i\} K_h^2(\mathbf{t}_i - \mathbf{t}) f(\mathbf{t}_i) d\mathbf{t}_i \\
= & \begin{pmatrix} \frac{\exp[H\{\eta(\mathbf{t})\}]}{(1+\exp[H\{\eta(\mathbf{t})\}])^2} f(\mathbf{t}) h^{-d} \int K^2(\mathbf{u}) d\mathbf{u} & \frac{\exp[H\{\eta(\mathbf{t})\}]}{(1+\exp[H\{\eta(\mathbf{t})\}])^2} f(\mathbf{t}) h^{m-1-d} \int K^2(\mathbf{u}) d\mathbf{u} \\ \frac{\exp[H\{\eta(\mathbf{t})\}]}{(1+\exp[H\{\eta(\mathbf{t})\}])^2} f(\mathbf{t}) h^{m-1-d} \int K^2(\mathbf{u}) d\mathbf{u} & \frac{\exp[H\{\eta(\mathbf{t})\}]}{(1+\exp[H\{\eta(\mathbf{t})\}])^2} f(\mathbf{t}) h^{2m-2-d} \int K^2(\mathbf{u}) d\mathbf{u} \end{pmatrix}.
\end{aligned}
$$

Thus $\mathrm{var}(\widehat{b}_0) = O\{(nh^d)^{-1}\}$ and $\mathrm{var}(\widehat{\mathbf{b}}_1) = O\{(nh^{d+2})^{-1}\}$.

The efficient estimator $\widehat{\mathcal{B}}_2$ satisfies

$$
\begin{aligned}
\mathbf{0} &= n^{-1/2} \sum_{i=1}^n \mathbf{S}_{\mathrm{eff}}(T_i, \mathbf{x}_i, \widehat{\mathcal{B}}_2^{\mathrm{T}} \mathbf{x}_i, \widehat{\eta}, \widehat{\boldsymbol{\eta}}', \widehat{E}) \\
&= n^{-1/2} \sum_{i=1}^n \mathrm{vecl}\left(\{\mathbf{x}_i - \widehat{E}(\mathbf{X}_i \mid \widehat{\mathcal{B}}_2^{\mathrm{T}} \mathbf{x}_i)\} \left[y_i - \frac{\exp\{\widehat{\eta}(\widehat{\mathcal{B}}_2^{\mathrm{T}} \mathbf{x}_i)\}}{1 + \exp\{\widehat{\eta}(\widehat{\mathcal{B}}_2^{\mathrm{T}} \mathbf{x}_i)\}} \right] \widehat{\boldsymbol{\eta}}'(\widehat{\mathcal{B}}_2^{\mathrm{T}} \mathbf{x}_i)^{\mathrm{T}} \right) \\
&= n^{-1/2} \sum_{i=1}^n \mathrm{vecl}\left(\{\mathbf{x}_i - \widehat{E}(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}} \mathbf{x}_i)\} \left[T_i - \frac{\exp\{\widehat{\eta}(\mathcal{B}^{\mathrm{T}} \mathbf{x}_i)\}}{1 + \exp\{\widehat{\eta}(\mathcal{B}^{\mathrm{T}} \mathbf{x}_i)\}} \right] \widehat{\boldsymbol{\eta}}'(\mathcal{B}^{\mathrm{T}} \mathbf{x}_i)^{\mathrm{T}} \right) \\
&\quad + n^{-1} \sum_{i=1}^n \frac{\partial}{\partial(\mathrm{vecl}\ \mathcal{B})^{\mathrm{T}}} \mathrm{vecl}\left(\{\mathbf{x}_i - \widehat{E}(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}} \mathbf{x}_i)\} \left[T_i - \frac{\exp\{\widehat{\eta}(\mathcal{B}^{\mathrm{T}} \mathbf{x}_i)\}}{1 + \exp\{\widehat{\eta}(\mathcal{B}^{\mathrm{T}} \mathbf{x}_i)\}} \right] \widehat{\boldsymbol{\eta}}'(\mathcal{B}^{\mathrm{T}} \mathbf{x}_i)^{\mathrm{T}} \right) \Bigg|_{\mathcal{B} = \mathcal{B}^*} \\
&\quad \times \sqrt{n}\, \mathrm{vecl}(\widehat{\mathcal{B}}_2 - \mathcal{B}) \\
&= n^{-1/2} \sum_{i=1}^n \mathrm{vecl}\left(\{\mathbf{x}_i - \widehat{E}(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}} \mathbf{x}_i)\} \left[T_i - \frac{\exp\{\widehat{\eta}(\mathcal{B}^{\mathrm{T}} \mathbf{x}_i)\}}{1 + \exp\{\widehat{\eta}(\mathcal{B}^{\mathrm{T}} \mathbf{x}_i)\}} \right] \widehat{\boldsymbol{\eta}}'(\mathcal{B}^{\mathrm{T}} \mathbf{x}_i)^{\mathrm{T}} \right) \\
&\quad + E\left\{ \frac{\partial}{\partial(\mathrm{vecl}\ \mathcal{B})^{\mathrm{T}}} \mathrm{vecl}\left(\{\mathbf{X}_i - E(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}} \mathbf{X}_i)\} \left[T_i - \frac{\exp\{\eta(\mathcal{B}^{\mathrm{T}} \mathbf{X}_i)\}}{1 + \exp\{\eta(\mathcal{B}^{\mathrm{T}} \mathbf{X}_i)\}} \right] \boldsymbol{\eta}'(\mathcal{B}^{\mathrm{T}} \mathbf{X}_i)^{\mathrm{T}} \right) \right\} \\
&\quad \times \sqrt{n}\, \mathrm{vecl}(\widehat{\mathcal{B}}_2 - \mathcal{B}) + o_p(1) \\
&= n^{-1/2} \sum_{i=1}^n \mathrm{vecl}\left(\{\mathbf{x}_i - \widehat{E}(\mathbf{X}_i \mid \mathcal{B}^{\mathrm{T}} \mathbf{x}_i)\} \left[T_i - \frac{\exp\{\widehat{\eta}(\mathcal{B}^{\mathrm{T}} \mathbf{x}_i)\}}{1 + \exp\{\widehat{\eta}(\mathcal{B}^{\mathrm{T}} \mathbf{x}_i)\}} \right] \widehat{\boldsymbol{\eta}}'(\mathcal{B}^{\mathrm{T}} \mathbf{x}_i)^{\mathrm{T}} \right) \\
&\quad - \mathbf{V} \sqrt{n}\, \mathrm{vecl}(\widehat{\mathcal{B}}_2 - \mathcal{B}) + o_p(1),
\end{aligned}
$$

where the last equality is because of the derivative of the efficient score is its negative second moment. We now expand the first term above as

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\mathrm{vecl}\left(\{\mathbf{x}_i-\widehat{E}(\mathbf{X}_i\mid\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i-\frac{\exp\{\widehat{\eta}(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\widehat{\eta}(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\widehat{\boldsymbol{\eta}}'(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)\\
= & n^{-1/2}\sum_{i=1}^{n}\mathrm{vecl}\left(\{\mathbf{x}_i-E(\mathbf{X}_i\mid\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i-\frac{\exp\{\eta(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\boldsymbol{\eta}'(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)\\
& +n^{-1/2}\sum_{i=1}^{n}\mathrm{vecl}\left(\{\mathbf{x}_i-E(\mathbf{X}_i\mid\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i-\frac{\exp\{\eta(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\{\widehat{\boldsymbol{\eta}}'(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}-\boldsymbol{\eta}'(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\}\right)\\
& -n^{-1/2}\sum_{i=1}^{n}\mathrm{vecl}\left(\{\mathbf{x}_i-E(\mathbf{X}_i\mid\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[\frac{\exp\{\widehat{\eta}(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\widehat{\eta}(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}-\frac{\exp\{\eta(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\widehat{\boldsymbol{\eta}}'(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)\\
& +n^{-1/2}\sum_{i=1}^{n}\mathrm{vecl}\left(\{\widehat{E}(\mathbf{X}_i\mid\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)-E(\mathbf{X}_i\mid\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[T_i-\frac{\exp\{\eta(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\widehat{\boldsymbol{\eta}}'(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right)\\
& +n^{-1/2}\sum_{i=1}^{n}\mathrm{vecl}\left(\{\widehat{E}(\mathbf{X}_i\mid\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)-E(\mathbf{X}_i\mid\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}\left[\frac{\exp\{\eta(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\eta(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}-\frac{\exp\{\widehat{\eta}(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}{1+\exp\{\widehat{\eta}(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)\}}\right]\right.\\
& \left.\times\widehat{\boldsymbol{\eta}}'(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right).
\end{aligned}
$$

It is easy to check that the second term is of order $o_p(1)$ since $\widehat{\boldsymbol{\eta}}'(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}-\boldsymbol{\eta}'(\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}=o_p(1)$ and the multiplication term in front has mean zero conditional on $\mathcal{B}^{\mathrm{T}}\mathbf{x}_i$. Similarly, the third term is of order $o_p(1)$ since $\mathbf{X}_i-E(\mathbf{X}_i\mid\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i)$ has conditional mean zero and is multiplied by an $o_p(1)$ term. The fourth term is of oder $o_p(1)$ for the same reason. Finally, the last term is trivially $o_p(1)$. We thus have obtained

$$
\mathbf{0}=n^{-1/2}\sum_{i=1}^{n}\mathbf{S}_{\mathrm{eff}}(T_i,\mathbf{x}_i,\ \mathcal{B}^{\mathrm{T}}\mathbf{x}_i,\eta,\boldsymbol{\eta}',E)-\mathbf{V}\sqrt{n}\mathrm{vecl}(\widehat{\mathcal{B}}_2-\ \mathcal{B})+o_p(1),
$$

hence the theorem follows. □

S.4 Proof of Lemma 3

Write $\widehat{\pi}(\mathbf{X}_i) = \pi(\mathbf{X}_i, \widehat{\boldsymbol{\gamma}})$. From (2.9), we write

$$
\begin{aligned}
& n^{1/2}(\widehat{\tau} - \tau) \\
= \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{Y_i T_i}{\pi(\mathbf{X}_i)} - \frac{Y_i(1-T_i)}{1-\pi(\mathbf{X}_i)} - \tau \right\} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{Y_i T_i}{\widehat{\pi}(\mathbf{X}_i)} - \frac{Y_i T_i}{\pi(\mathbf{X}_i)} - \frac{Y_i(1-T_i)}{1-\widehat{\pi}(\mathbf{X}_i)} + \frac{Y_i(1-T_i)}{1-\pi(\mathbf{X}_i)} \right\} \\
= \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{Y_i T_i}{\pi(\mathbf{X}_i)} - \frac{Y_i(1-T_i)}{1-\pi(\mathbf{X}_i)} - \tau \right\} \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{Y_i T_i}{\pi^2(\mathbf{X}_i)} \{\pi(\mathbf{X}_i) - \widehat{\pi}(\mathbf{X}_i)\} - \frac{Y_i(1-T_i)}{\{1-\pi(\mathbf{X}_i)\}^2} \{\widehat{\pi}(\mathbf{X}_i) - \pi(\mathbf{X}_i)\} \right] \\
& + O_p\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\widehat{\pi}(\mathbf{X}_i) - \pi(\mathbf{X}_i)\}^2 \right] \\
= \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{Y_i T_i}{\pi(\mathbf{X}_i)} - \frac{Y_i(1-T_i)}{1-\pi(\mathbf{X}_i)} - \tau \right\} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{Y_i T_i}{\pi^2(\mathbf{X}_i)} + \frac{Y_i(1-T_i)}{\{1-\pi(\mathbf{X}_i)\}^2} \right] \frac{\partial \pi(\mathbf{X}_i, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_0} (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \\
& + O_p(n^{-1/2}) \\
= \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{Y_i T_i}{\pi(\mathbf{X}_i)} - \frac{Y_i(1-T_i)}{1-\pi(\mathbf{X}_i)} - \tau \right\} \\
& - E\left(\left[\frac{Y_i T_i}{\pi^2(\mathbf{X}_i)} + \frac{Y_i(1-T_i)}{\{1-\pi(\mathbf{X}_i)\}^2} \right] \frac{\partial \pi(\mathbf{X}_i, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_0} \right) n^{-1/2} \sum_{i=1}^{n} \phi(\mathbf{X}_i, T_i) + o_p(1) \\
= \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{Y_i T_i}{\pi(\mathbf{X}_i)} - \frac{Y_i(1-T_i)}{1-\pi(\mathbf{X}_i)} - \tau \right\} \\
& - E\left(\left[\frac{Y_i^*(1)}{\pi(\mathbf{X}_i)} + \frac{Y_i^*(0)}{1-\pi(\mathbf{X}_i)} \right] \frac{\partial \pi(\mathbf{X}_i, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_0} \right) n^{-1/2} \sum_{i=1}^{n} \phi(\mathbf{X}_i, T_i) + o_p(1) \\
= \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\left\{ \frac{Y_i T_i}{\pi(\mathbf{X}_i)} - \frac{Y_i(1-T_i)}{1-\pi(\mathbf{X}_i)} - \tau \right\} - \left\{ \frac{Y_i^*(1)}{\pi(\mathbf{X}_i)} + \frac{Y_i^*(0)}{1-\pi(\mathbf{X}_i)} \right\} \{T_i - \pi(\mathbf{X}_i)\} \right] \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\left\{ \frac{Y_i^*(1)}{\pi(\mathbf{X}_i)} + \frac{Y_i^*(0)}{1-\pi(\mathbf{X}_i)} \right\} \{T_i - \pi(\mathbf{X}_i)\} - E\left(\left[\frac{Y_i^*(1)}{\pi(\mathbf{X}_i)} + \frac{Y_i^*(0)}{1-\pi(\mathbf{X}_i)} \right] \frac{\partial \pi(\mathbf{X}_i, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_0} \right) \phi(\mathbf{X}_i, T_i) \right] \\
& + o_p(1) \\
= \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\left\{ \frac{Y_i T_i}{\pi(\mathbf{X}_i)} - \frac{Y_i(1-T_i)}{1-\pi(\mathbf{X}_i)} - \tau \right\} - \left\{ \frac{Y_i^*(1)}{\pi(\mathbf{X}_i)} + \frac{Y_i^*(0)}{1-\pi(\mathbf{X}_i)} \right\} \{T_i - \pi(\mathbf{X}_i)\} \right] \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} B_i + o_p(1).
\end{aligned}
$$

Comparing with the results in Hirano et al. (2003), it is now clear that the first line at the last expression above is the efficient influence function, while the second line is the difference between the influence functions of the estimator based on the parametric estimator of π and the efficient estimator, hence is orthogonal to the efficient influence function. In fact the orthogonality is also easily checked by direct calculation. This difference is the price we have to pay when we use a parametric model to estimate π instead of doing it fully nonparametrically. $\square$

S.1 A simulation study on estimating the average treatment effect

We generate the potential outcomes as follows: $Y^*(1) = |X_1 X_5| + \sin\left(\sum_{j=1}^{6} X_j\right)$ and $Y^*(0) = 0$. We generate the treatment indicator following (2.2) while considering the two different functional forms for $\eta(\cdot)$ as specified in Section 4.1 .

We compare estimating the average treatment effect using formula (2.9) but different methods to estimate the propensity function: "True" (use the true form of the propensity score function with known parameters), "Oracle" (use the true form of the propensity score function but estimate the unknown parameters), "Logistic" (use a logistic regression model to estimate the propensity score function), $\hat{\tau}_1$ (proposed semiparametric estimator with $\mathcal{B}$ being estimated by $\widehat{\mathcal{B}}_1$), and $\hat{\tau}_2$ (proposed semiparametric estimator with $\mathcal{B}$ being estimated by $\widehat{\mathcal{B}}_2$). The boxplots of the estimated average treatment effect (based on 1000 simulation runs) using these five methods are displayed in Figure S1. We observe that the "Logistic" method based on a misspecified propensity score function has serious bias; while the performance of the proposed semiparametric estimators are close to that of "True" and "Oracle". Note that while the "Oracle" estimator is the gold standard in terms of

$\mathcal{B}$ and propensity score estimation in Section 4.1, it is unclear that it should yield the best treatment effect estimation here. Hence we included it as a "standard result" for completeness.

S.2 Comparison with Other Methods

We compare our semiparametric approach with Tan's improved methods (Tan) (Tan, 2006, 2010), targeted maximum likelihood estimation (TMLE) (van der Laan and Rubin, 2006) and the biased reduced double robust (BRdr) estimator proposed by Vermeulen and Vansteelandt (2015). Because Tan's method requires implementing a regression model on treatment outcome $Y^*(1), Y^*(0)$ separately, we slightly modified $Y^*(0)$ to follow $N(0,1)$ in order to implement the method. We summarize the average treatment effect in Figure (S2) and Figure (S3).

We then consider the case where the true outcome model is indeed a linear model. Specifically, we set $Y = X_1 + X_2 + TX_3 + X_4 + 13.5X_5 + X_6 + \epsilon, \epsilon \sim \text{Normal}(0, 5^2)$ when $d = 1$ and let $Y = X_1 + X2 + X_3 + X_4 + 3.5X_5 + TX_6 + \epsilon, \epsilon \sim \text{Normal}(0, 5^2)$ when $d = 2$. We compare the average treatment effect estimates in Figure (S4) and Figure (S5), respectively.

References

Hirano, K., Imbens, G. W., and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. Econometrica **71,** 1161–1189.

Tan, Z. (2006). A distributional approach for causal inference using propensity scores. Journal of the American Statistical Association **101,** 1619–1637.

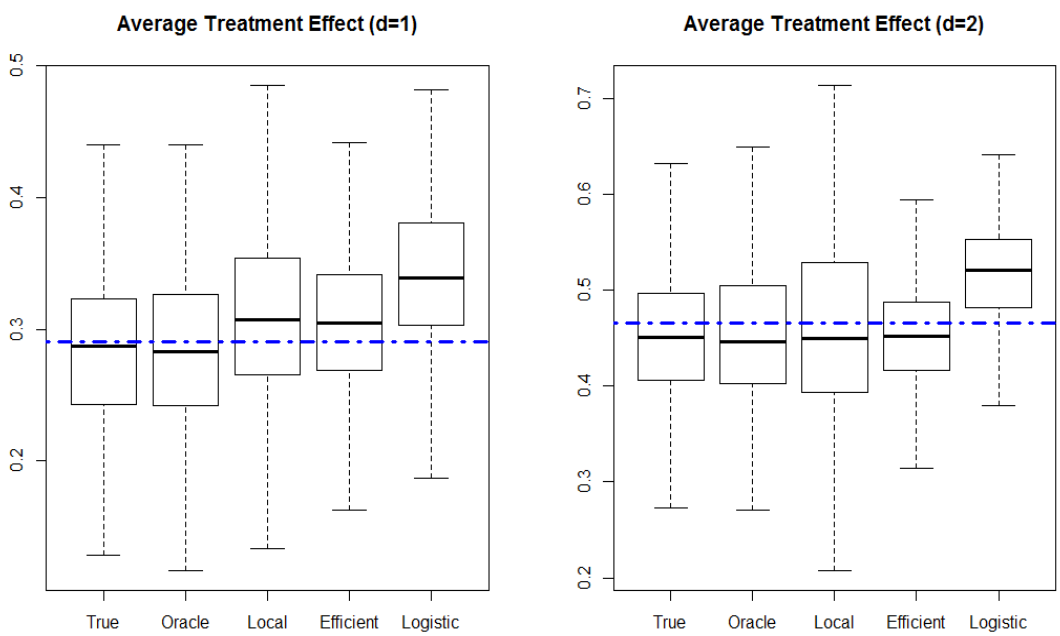

Figure S1: Average treatment effect in example 1 (left) and example 2 (right). The blue dash line is the true average treatment effect.

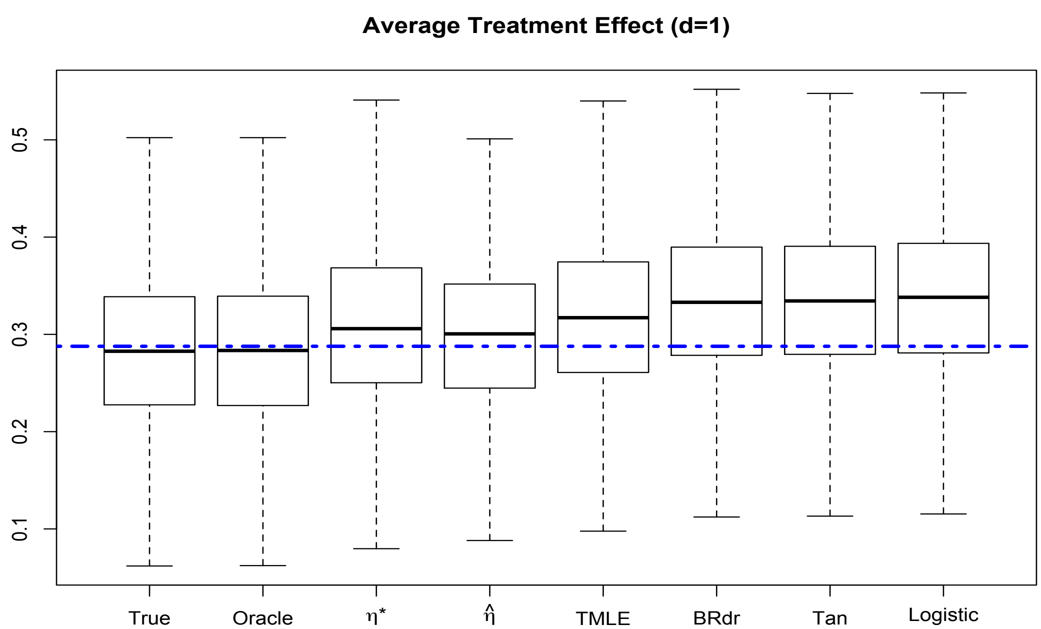

Figure S2: Average treatment effect in example 1. The blue dashed line is the true average treatment effect.

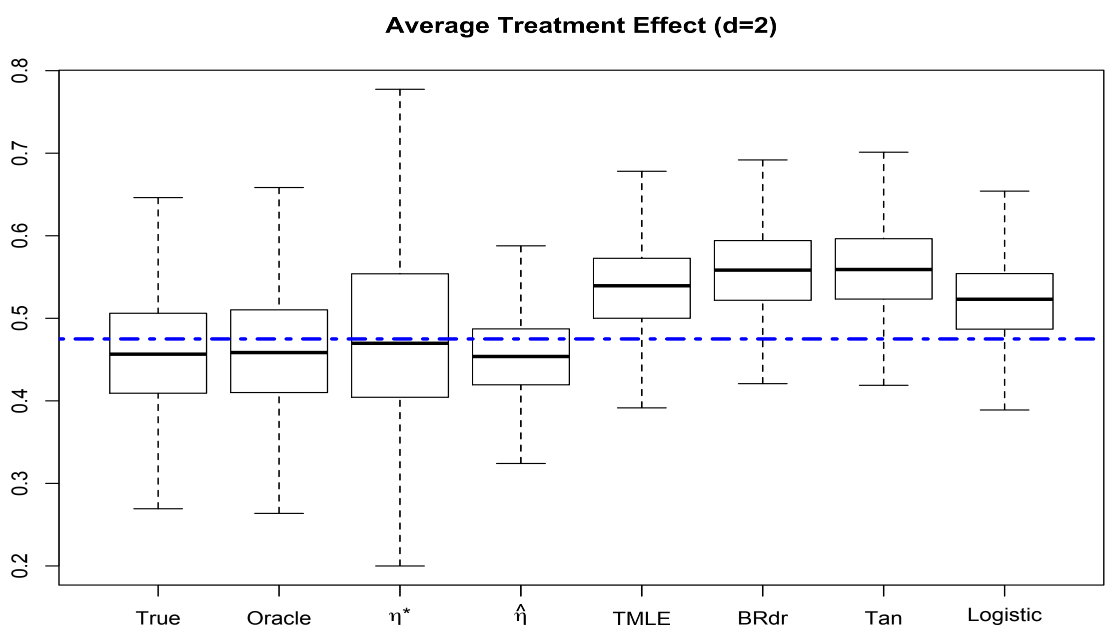

Figure S3: Average treatment effect in example 2. The blue dash line is the true average treatment effect.

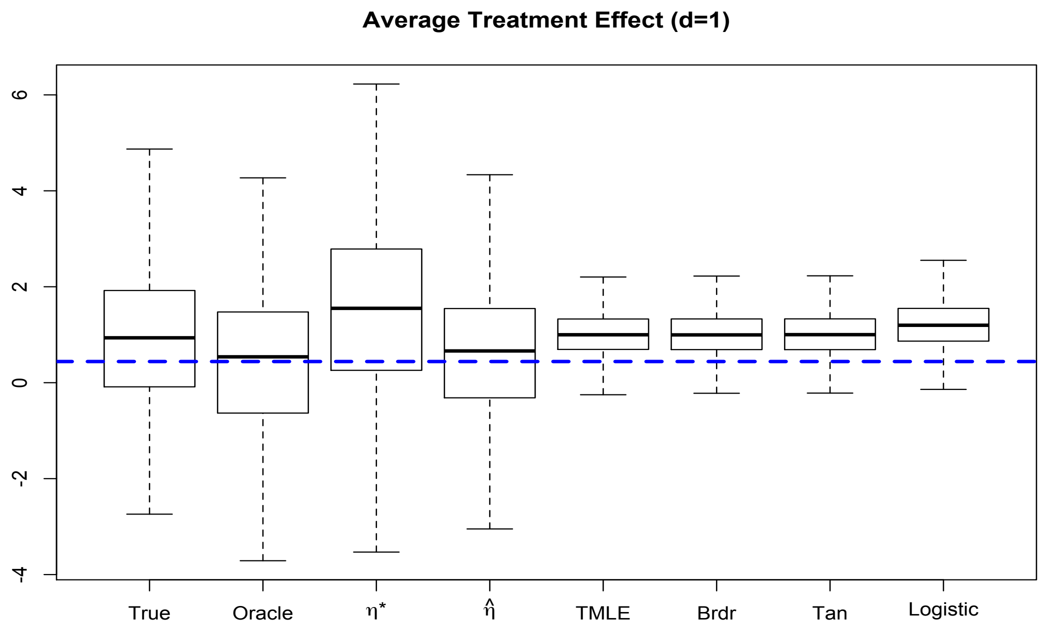

Figure S4: Average treatment effect in example 1, where the outcome is $Y = X_1 + X_2 + TX_3 + X_4 + 13.5X_5 + X_6 + \epsilon, \epsilon \sim \mathrm{N}ormal(0, 5^2)$. The blue dash line is the true average treatment effect.

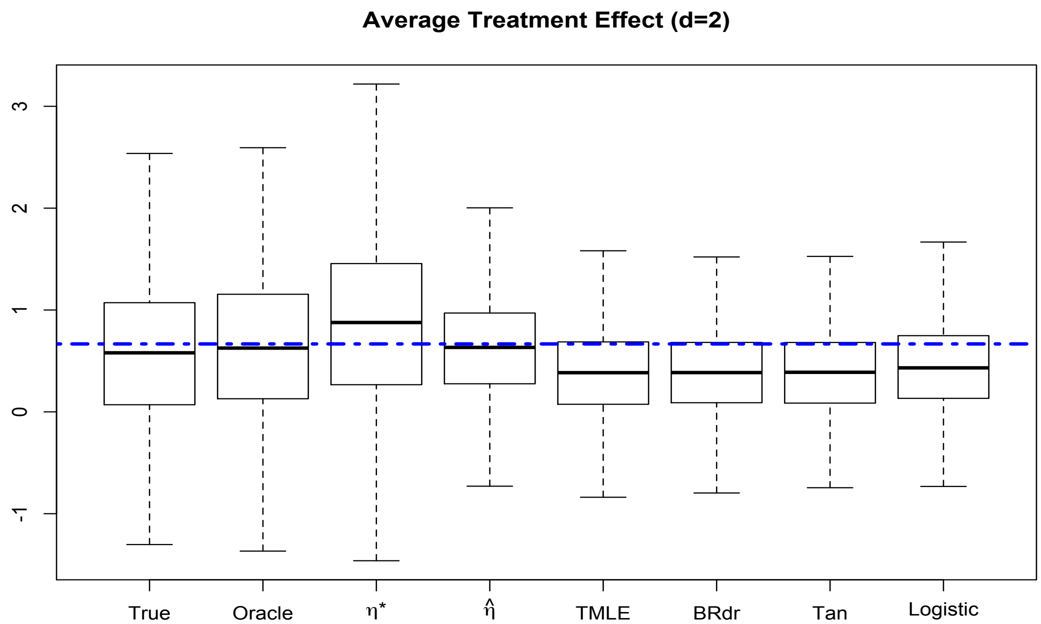

Figure S5: Average treatment effect in example 2, where the outcome is $Y = X_1 + X2 + X_3 + X_4 + 3.5X_5 + TX_6 + \epsilon, \epsilon \sim \mathrm{N}ormal(0, 5^2)$. The blue dash line is the true average treatment effect.

Tan, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. Biometrika **97,** 661–682.

van der Laan, M. J. and Rubin, D. (2006). Targeted maximum likelihood learning. The International Journal of Biostatistics **2,** 1–40.

Vermeulen, K. and Vansteelandt, S. (2015). Bias-reduced doubly robust estimation. Journal of the American Statistical Association **110,** 1024–1036.